import pandas as pd

import numpy as np

import collections

import re

print("All Modules Imported")

#Sample documents

doc1 = 'Game of Thrones is an amazing tv series!, Game of Thrones is the best tv series! and Game of Thrones is so great'

#Sentance without punctuations and split them

w\_doc1= re.sub(r'[^\w\s\\_]','', doc1.lower()).split()

# Print the sentence without punctuation

print(w\_doc1)

import nltk

from nltk.corpus import stopwords

nltk.download('stopwords')

stop\_words = set(stopwords.words('english'))

filtered\_words = [word for word in w\_doc1 if word.lower() not in stop\_words]

# Reconstruct the text without stop words

filtered\_text = ' '.join(filtered\_words)

# Print the text without stop words

print(filtered\_text)

from sklearn.feature\_extraction.text import CountVectorizer

doc1 = ['Game of Thrones is an amazing tv series!, Game of Thrones is the best tv series! and Game of Thrones is so great']

# Create an instance of CountVectorizer

vectorizer = CountVectorizer()

# Fit the vectorizer on the sentences and transform them into a Bag of Words representation

X = vectorizer.fit\_transform(doc1)

# Get the feature names (words)

feature\_names = vectorizer.get\_feature\_names\_out()

# Convert the Bag of Words representation to a dense matrix and print it

print(X.toarray())

print("Feature names (words):", feature\_names)